
AI / ML Development

Task #1

Title: Fraud Detection System with Machine Learning

Scenario:

You are part of a team at a financial institution tasked with developing a fraud detection system. Your objective is to build a machine learning model that can accurately identify fraudulent transactions in real-time, helping the company minimize financial losses and maintain the trust of its customers.

Task Description:

Your task involves the following steps:

1. Data Collection and Exploration:

Obtain a dataset of transaction records, which includes features such as transaction amount, merchant ID, transaction time, etc. The dataset should contain a mix of both legitimate and fraudulent transactions. Explore the dataset to gain insights into its structure and distributions.

2. Feature Engineering:

Conduct feature engineering to extract relevant features from the raw transaction data. This may include creating new features, encoding categorical variables, handling missing values, and scaling numerical features.

3. Model Selection and Training:

Select appropriate machine learning algorithms for fraud detection, such as logistic regression, random forest, or gradient boosting classifiers. Train multiple models using the preprocessed data and evaluate their performance using appropriate metrics like precision, recall, and F1-score.

4. Imbalanced Data Handling:

Since fraudulent transactions are typically rare compared to legitimate ones, address the issue of class imbalance in the dataset. Implement techniques such as oversampling, undersampling, or synthetic data generation to balance the classes and improve the model's ability to detect fraud.

5. Hyperparameter Tuning:

Fine-tune the hyperparameters of the selected model(s) using techniques like grid search or random search to optimize their performance. Experiment with different parameter combinations to find the best configuration for detecting fraud while minimizing false positives.

6. Model Evaluation and Deployment:

Evaluate the final model(s) on a holdout dataset or using cross-validation to assess their generalization performance. Once satisfied with the model's performance, deploy it as a real-time fraud detection system that can analyze incoming transactions and flag suspicious ones for further review.

Submission Requirements:

- Organize your code into well-documented Python scripts or Jupyter notebooks. Include comments to explain key steps and decisions made during model development.
- Create a virtual environment for your project and perform all coding within this environment. Submit the entire virtual environment folder along with your code for evaluation. This ensures that the evaluation team can activate the same environment and run your code seamlessly.
- Provide a README file with clear instructions on how to activate the virtual environment, run the code, install dependencies, and reproduce the results. Include any additional notes or considerations for reviewers.
- Write a brief report summarizing your approach, including details on data collection, preprocessing steps, model architecture, training process, evaluation metrics, and any challenges faced. Discuss the implications of your findings and potential future directions for improvement.

Deadline:

The deadline for completing tasks and submitting final deliverables is June 3, 2024, allowing ample time for a thorough project completion.

Submission Process for Internship Tasks

Interns are required to submit their completed tasks via email. Please follow the instructions below to ensure your submission is correctly formatted and complete.

1. Email Submission:
 - Send an email to: submission@xpacetechnologies.com
2. Email Subject Format:
 - Use the following format for the subject of your email: ID+TaskName
 - For example: "AIMLINT-240324-XTxxxx+FraudDetectionSystem"
3. Email Body:
 - Include a brief introduction of yourself and a summary of the project. Mention your full name, internship ID, and a short description of the task you are submitting.
4. Attachments:
 - Attach a PDF file containing the project documentation.

Example Email:

To: submission@xpacetechnologies.com

Subject: AIMLINT-240324-XTxxxx+AI | ML Development+FraudDetectionSystem

Body:

Dear Evaluation Team,

My name is [Your Name], and I am one of the interns participating in the AI/ML virtual internship program. I am pleased to submit my completed task for the Fraud Detection System project.

Brief Introduction:

I am [Your Name], currently pursuing [Your Degree] at [Your University]. I have a strong interest in machine learning and data science, and I have thoroughly enjoyed working on this project. The objective of the project was to develop a fraud detection system using machine learning techniques to identify fraudulent transactions in real-time.

Project Summary:

The project involved data collection and exploration, feature engineering, model selection and training, handling imbalanced data, hyperparameter tuning, and model evaluation and deployment. I have documented my entire process and findings in the attached PDF.

Attachments:

- Project Documentation (PDF file)

Thank you for this opportunity and I look forward to your feedback.

Best regards,

[Your Name]

[Your Contact Information]